BU-262-M

A MULTIVARIATE TEST OF SIGNIFICANCE FOR RESPONSES FROM ANY POPULATION I: UNRESTRICTED RANDOMIZATION[1/]

by

N. S. Urquhart[2/] and J. S. Williams

Cornell University and Colorado State University

Abstract

Many tests for mean differences utilize statistics based on a comparison of 'hypothesis' and 'residual' sums of squares. The probability distribution of such statistics usually is derived under the assumption of normally distributed errors. The frequent occurrence of data from non-normal distributions leads one to ask how closely the actual distributions are approximated by their normally derived counterparts. Permutation theory, which has been used to investigate this question in univariate situations, is applied here to the multivariate trace statistic. The results, summarized at the end of section 4, indicate that the permutation distribution of the trace statistic is reasonably approximated by its normally derived counterpart for a wide class of non-normal situations. For example, the nature of the regression matrix is an important determinant of this agreement. Otherwise an approximation incorporating fourth order k-statistics is suggested.

1/ Paper No. Bu-156 of the Biometrics Unit and No. 549 of the Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.

2/ Part of a Ph. D. thesis completed at Colorado State University by the senior author, June 1965. Research supported in part by a grant from the National Institutes of Health.

A MULTIVARIATE TEST OF SIGNIFICANCE FOR RESPONSES FROM ANY POPULATION I: UNRESTRICTED RANDOMIZATION[1/]

by

N. S. Urquhart[2/] and J. S. Williams

Cornell University and Colorado State University

Summary

Many tests for mean differences utilize statistics based on a comparison of 'hypothesis' and 'residual' sums of squares. The probability distribution of such statistics usually is derived under the assumption of normally distributed errors. The frequent occurrence of data from non-normal distributions leads one to ask how closely the actual distributions are approximated by their normally derived counterparts. Permutation theory, which has been used to investigate this question in univariate situations, is applied here to the multivariate trace statistic. The results, summarized at the end of section 4, indicate that the permutation distribution of the trace statistic is reasonably approximated by its normally derived counterpart for a wide class of non-normal situations. For example, the nature of the regression matrix is an important determinant of this agreement. Otherwise an approximation incorporating fourth order k-statistics is suggested.

1. Introduction

Fisher [1935] noted that the random assignment of treatments to experimental units suffices, without any distributional assumptions, to provide the distribution for an exact test of significance of differences among treatment means. He noted

---

1/ Paper No. BU-156 of the Biometrics Unit and No. 549 of the Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.

2/ Part of a Ph. D. thesis completed at Colorado State University by the senior author, June 1965. Research supported in part by a grant from the National Institutes of Health.

that the observed association of treatments with experimental units (and consequently with responses) is one of many which could occur under the randomization scheme. If there is no treatment difference, the response for each other association can be obtained merely by permuting treatment labels on the original association.

A statistic, for example the t-statistic in the simplest univariate case, can be evaluated for each permutation. Since each permutation is equally probable under the randomization scheme, the probability distribution of the statistic can be built up from the observed data under the hypothesis of no treatment effect. This "permutation distribution" can then be used to evaluate the significance of the observed outcome.

Various workers have applied this approach to study specific test statistics. Pitman [1937] used Fisher's idea to study the permutation distribution of a form of the usual treatment F-test for a univariate response from a randomized block design. He was able to show that the normally derived distribution served as a satisfactory approximation to the permutation distribution provided adjustments based on the data were incorporated into the standard analysis. Welch [1937] encountered algebraic problems in a similar study of latin square designs because the randomization is over the class of latin squares rather than over plots within a block. However his conclusions were about the same as Pitman's. Arnold [1964] investigated the paired comparison Hotelling's T^2 with permutation theory. He also found reasonable agreement between the distribution based on normal theory and the exact permutation distribution provided adjustments based on the data were incorporated into the standard analysis.

Fisher's concept can be extended to regression-type experiments. For example, Box and Watson [1962] considered a univariate multiple regression model. They

derived the first two moments of $W = S_0/(S_0 + S_E)$ (S_0 and S_E are the hypothesis and error sums of squares, respectively), and noted the importance of the independent-variate configuration in determining the agreement between the moments derived under permutation theory and under normal theory.

The models discussed above exemplify two approaches: the recognition of influence of treatments on the population structure in contrast to the multiple regression approach. We wish to make a compromise between these approaches as well as consider a multivariate response. We will speak of treatments and randomization, but will use a general regression model for the algebraic development. Thus our discussion can be easily interpreted by the reader for regression problems we do not cover, but which fit within the multivariate generalization of Box and Watson's model.

In particular we are concerned with a further study of the role of the treatment (independent-variate) configuration in determining the agreement between permutation theory and normal theory. This raises three basic problems: (i) unrestricted randomization where choice of the number of replicates per treatment is the controllable element of the permutation distribution (Part I: Unrestricted Randomization); (ii) restricted randomization in blocks of equal size where numbers of replicates and design balance are the controllable elements of the permutation distribution (Part II: Restricted Randomization and Blocked Experimental Designs); and (iii) a study of the way in which the common experimental designs affect the permutation distribution (Part III: Performance in Common Experimental Designs).

This paper considers the first topic above. The other two topics will be the subjects of subsequent papers. This development is along the lines of Fisher's concept of an exact permutation test of significance and uses the methods of Pitman

and Box and Watson. The model considered here is a generalization of Box and Watson's model to a multivariate response. We consider $T = tr[\underset{\sim}{W}_H(\underset{\sim}{W}_H + \underset{\sim}{W}_E)^-]$ as a multivariate generalization of the beta form of the analysis of variance test criterion. The matrices $\underset{\sim}{W}_H$ and $\underset{\sim}{W}_E$ are the hypothesis and error sums of squares and cross products matrices, respectively, and $(\underset{\sim}{W}_H + \underset{\sim}{W}_E)^-$ denotes the generalized inverse of the matrix $\underset{\sim}{W}_H + \underset{\sim}{W}_E$ as defined, for example by Searle [1965]. The choice of T over other multivariate generalizations of $S_0/(S_0 + S_E)$ is justified on bases other than simplicity in the final section.

2. The Permutation Model

Experimenters frequently want to compare the effects of a set of treatments to see if they produce different responses from experimental material. An experiment may be run by randomly allocating t treatments to N experimental units. Treatments may be unrelated, as for example are varieties, in which case there should be more experimental units than treatments. Or the treatments may be related as are the various factor combinations of a factorial experiment. More generally the treatments may represent the application of different levels of several stimuli applied to the experimental units. There is no requirement that they constitute anything like a complete factorial set of treatments.

It is in this manner of defining a treatment that we are able to discuss standard treatment comparisons simultaneously with expected responses to varied levels of a set of factors. This practice has also become common in the context of the theory of linear models.

Each of these situations fits within the framework of a multivariate regression model which involves the two matrices $\underset{\sim}{Y}_{N\times q}$ and $\underset{\sim}{X}_{N\times p}$. The meaning of these matrices lies within the nature of the experiment and the underlying population structure. In order to clarify the role of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ we will discuss the population

structure, a model for the observed responses, the effect of randomization, and a test for the hypothesis of no treatment effect, in that order.

2.1 The Population Structure

Index the t treatments by $i(i=1,2,\cdots,t)$ and the N experimental units by $j(j=1,2,\cdots,N)$. If the i^{th} treatment were applied to each experimental unit a finite population of N q-variate responses $\underset{\sim}{w}_{ij}$ would result. Their population mean is given by

$$\underset{\sim}{\bar{w}}_{i\cdot} = \frac{1}{N} \sum_{j=1}^{N} \underset{\sim}{w}_{ij}$$

and the grand mean of all t of the populations is

$$\underset{\sim}{\bar{w}}_{\cdot\cdot} = \frac{1}{Nt} \sum_{i=1}^{t} \sum_{j=1}^{N} \underset{\sim}{w}_{ij} .$$

Of course the actual experiment produces only N of these Nt possible responses since the experimenter randomly allocates the t treatments to the N experimental units and thereby samples (or realizes) r_1 elements from $\{\underset{\sim}{w}_{1j}\}$, r_2 from $\{\underset{\sim}{w}_{2j}\}$, and so on with $r_1 + r_2 + \cdots + r_t = N$ and $r_i \geq 1$. The only restriction on this sampling is that a given experimental unit can yield a response to only one treatment. Consequently, the list of all possible samples corresponds to all possible permutations of the N plot indices into groups of size $r_1, r_2, \cdots, r_t$. This may be conceived as the association of plots with treatments or treatments with plots, whichever is more convenient.

2.2 A Model for the Observed Responses

In the completed experiment, the response from the j^{th} plot may be denoted by $\underset{\sim}{y}_{ik}$, $i=1,2,\cdots,t$, $k=1,2,\cdots,r_i$, i.e. $\underset{\sim}{y}_{ik}$ is the k^{th} response observed on treatment i. By comparing the definitions of $\underset{\sim}{y}_{ik}$ and $\underset{\sim}{w}_{ij}$, it follows that

$\{\underset{\sim}{y}_{ik} | k=1,2,\cdots,r_i\} \subset \{\underset{\sim}{w}_j | j=1,2,\cdots,N\}$. This observation is essential to the development of an inferential model which is additive and exact:

$$\underset{\sim}{y}_{ik} = \underset{\sim}{\bar{w}}_{i\cdot} + (\underset{\sim}{y}_{ik} - \underset{\sim}{\bar{w}}_{i\cdot})$$

$$= \underset{\sim}{\bar{w}}_{\cdot\cdot} + (\underset{\sim}{\bar{w}}_{i\cdot} - \underset{\sim}{\bar{w}}_{\cdot\cdot}) + (\underset{\sim}{y}_{ik} - \underset{\sim}{\bar{w}}_{i\cdot}) .$$

The rows of the matrix which we now denote by $\underset{\sim}{Y}$ will be these elements corrected for the sample mean response, i.e.

$$\underset{\sim}{y}'_{ik} - \underset{\sim}{\bar{y}}'_{\cdot\cdot} = (\underset{\sim}{\bar{w}}'_{i\cdot} - \underset{\sim}{\bar{w}}'_{\cdot\cdot}) + [(\underset{\sim}{y}'_{ik} - \underset{\sim}{\bar{y}}'_{\cdot\cdot}) - (\underset{\sim}{\bar{w}}'_{i\cdot} - \underset{\sim}{\bar{w}}'_{\cdot\cdot})], \qquad (2.1)$$

with the responses to the first treatment appearing in the first r_1 rows, to the next treatment in the next r_2 rows, and so on. The expression on the right-hand side of (2.1) may be regarded as the sum of treatment and residual effects. We note here that if $\underset{\sim}{j}_N$ is a column of N ones, the definition of $\underset{\sim}{Y}$ gives $\underset{\sim}{j}_N'\underset{\sim}{Y} = \underset{\sim}{0}$, the null vector.

The next step is to introduce the regression matrix, $\underset{\sim}{X}$, into the model. Each treatment may be characterized by the levels of each of the p factors common to one or more of the set of treatments, i.e. by $\underset{\sim}{x}_i' = (x_{i1},\cdots,x_{ip})$. The levels are taken as deviations from their mean values in the experiment. Then if $\underset{\sim}{X}_1 = (\underset{\sim}{x}_i')$ and if

$$\begin{bmatrix} \underset{\sim}{\bar{w}}_{1\cdot} - \underset{\sim}{\bar{w}}_{\cdot\cdot} \\ \underset{\sim}{\bar{w}}_{2\cdot} - \underset{\sim}{\bar{w}}_{\cdot\cdot} \\ \vdots \\ \underset{\sim}{\bar{w}}_{t\cdot} - \underset{\sim}{\bar{w}}_{\cdot\cdot} \end{bmatrix} = \underset{\sim}{X}_1\underset{\sim}{B} \qquad (2.2)$$

has a solution for $\underset{\sim}{B}$, we can express (2.1) as

$$\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{x}_1' \otimes \underset{\sim}{j}_{r_1} \\ \vdots \\ \underset{\sim}{x}_t' \otimes \underset{\sim}{j}_{r_t} \end{bmatrix} + \begin{bmatrix} (\underset{\sim}{y}_{11} - \bar{\underset{\sim}{y}}..) - (\bar{\underset{\sim}{w}}_1. - \bar{\underset{\sim}{w}}..) \\ \vdots \\ (\underset{\sim}{y}_{t r_t} - \bar{\underset{\sim}{y}}..) - (\bar{\underset{\sim}{w}}_t. - \bar{\underset{\sim}{w}}..) \end{bmatrix}$$

$$= \underset{\sim\sim}{X}\underset{\sim\sim}{B} + \underset{\sim}{R} \text{ , say.} \qquad (2.3)$$

Since the x's were taken as deviations from their respective experiment means, $\underset{\sim}{j}_N'\underset{\sim\sim}{X} = \underset{\sim}{0}$.

If (2.2) does not have a solution, it is possible to reason heuristically that (2.3) should still be applicable. The experimenter includes matrix $\underset{\sim}{X}$ in the formulation since he feels that it may help explain variation in the values of $\underset{\sim}{Y}$. It could be used in many ways, but the simplest is to hope that some linear combination, say $\underset{\sim\sim}{X}\underset{\sim\sim}{B}$, serves as a prediction for $\underset{\sim}{Y}$. That is, the relation (2.3) is assumed to be true, but now $\underset{\sim}{R}$ is defined by $\underset{\sim}{R} = \underset{\sim}{Y} - \underset{\sim\sim}{X}\underset{\sim\sim}{B}$ rather than by the identity (2.3).

2.3 The Effect of Randomization

There are $N!/(r_1!\, r_2! \cdots r_t!)$ possible experimental outcomes which have equal probability of occurring under the randomization scheme. One of these is given by the observed outcome. The other possible outcomes can be related to the model (2.3). Any other outcome would represent responses to the same treatments, but applied to other experimental units. Let us agree to record the responses in the same plot order as $\underset{\sim}{Y}$ in (2.3) so that a different sample would involve one or more interchanges of rows between partitions in $\underset{\sim}{X}$. This is easily accomplished by premultiplying the matrix $\underset{\sim}{X}$ by a permutation matrix $\underset{\sim}{P}$, i.e. $\underset{\sim\sim}{P}\underset{\sim\sim}{X}$ replaces $\underset{\sim}{X}$.

The outcome corresponding to the treatment order $\underset{\sim\sim}{P}\underset{\sim\sim}{X}$ will have responses which

we denote by $\underset{\sim}{Y}_p$. If the plot residuals, that is the rows of $\underset{\sim}{R}$, do not depend on the treatment applied, then the outcome corresponding to the treatment order $\underset{\sim}{P}\underset{\sim}{X}$ is modeled by

$$\underset{\sim}{Y}_p = \underset{\sim}{P}\underset{\sim}{X}\underset{\sim}{B} + \underset{\sim}{R} . \tag{2.4}$$

This is an association of treatment labels with responses in plot order. However since $\underset{\sim}{P}$ is an orthogonal matrix

$$\underset{\sim}{P}'\underset{\sim}{Y}_p = \underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{X}\underset{\sim}{B} + \underset{\sim}{P}'\underset{\sim}{R}$$

$$= \underset{\sim}{X}\underset{\sim}{B} + \underset{\sim}{R}$$

which is an association of plot responses with treatments in a standard order.

2.4 A Test for the Hypothesis of No Treatment Effect

In the proposed absence of treatment effects upon which significance will be calculated, the observed yields constitute a uniformity trial, that is $\{\underset{\sim}{w}_{1j}\} = \{\underset{\sim}{w}_{2j}\} = \cdots = \{\underset{\sim}{w}_{tj}\}$ and consequently $\underset{\sim}{X}\underset{\sim}{B} = \underset{\sim}{0}$. From (2.4) then we see that $\underset{\sim}{Y}_p = \underset{\sim}{R} = \underset{\sim}{Y}$; the coupling of $\underset{\sim}{Y}_p$ with $\underset{\sim}{P}\underset{\sim}{X}$ manifests only the random assignment of treatments to experimental units, i.e. $(\underset{\sim}{Y}_p, \underset{\sim}{P}\underset{\sim}{X}) = (\underset{\sim}{Y}, \underset{\sim}{P}\underset{\sim}{X})$. All such pairs are equally probable.

The absence of treatment effects should be tested with a statistic sensitive to departures from the null hypothesis. The class of tests based on $\underset{\sim}{W}_H = \underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}$ and $\underset{\sim}{W}_E = \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{W}_H$ should possess more than a minimal element of sensitivity to deviations from H_0: $\underset{\sim}{X}\underset{\sim}{B} = \underset{\sim}{0}$ since $\underset{\sim}{W}_H$ and $\underset{\sim}{W}_E$ both will be proportional to unbiased estimates of the plot residual dispersion matrix only if the null hypothesis is true for the basis set of variates in Y. Further, $\underset{\sim}{W}_H$ and $\underset{\sim}{W}_E$ form the basis of nearly optimal tests when the plot residual variation is nearly normal. Within this class we consider

$$T = \mathrm{tr}[\underset{\sim}{W}_H(\underset{\sim}{W}_H + \underset{\sim}{W}_E)^{-}] = \mathrm{tr}[\underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}] = \mathrm{tr}[\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}\underset{\sim}{Y}'] \qquad (2.5)$$

as a potentially robust candidate for a test of the hypothesis of no treatment effect. In the last form of (2.5), $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ occur in a symmetric fashion in T. This symmetry forms the basis for our subsequent discussion of how "normality" of $\underset{\sim}{X}$ can substitute for normality of $\underset{\sim}{Y}$ and of the potentially robust character of T relative to other statistics based on $\underset{\sim}{W}_H$ and $\underset{\sim}{W}_E$.

The randomness in the value assumed by T (or any other statistic based on $\underset{\sim}{X}$ and $\underset{\sim}{Y}$) is a consequence of the random assignment of treatments to experimental units. Given $\underset{\sim}{Y}$ and $\underset{\sim}{X}$, the probability distribution of T, or certain properties of this distribution, e.g. its mean and variance, can be evaluated by recourse to this original source of randomness.

Extensive calculations are required for evaluating the permutation distribution of T. Following Pitman [1937] we approximate the permutation distribution by a beta distribution with the same first two moments. The use of such a beta approximation also has been employed in the normal theory of errors by Pillai [1955] to find an approximation to the distribution of T.

3. The Permutation Moments of T

The mean and variance of T are evaluated by averaging the values of T and T^2 for each possible association of treatments with experimental units. Recall that there are $N!/(r_1!\, r_2! \cdots r_t!)$ possible and equally probable associations. Thus each association has probability

$$\pi = \frac{r_1!\, r_2! \cdots r_t!}{N!} = (r_1!\, r_2! \cdots r_t!)\left(\frac{1}{N!}\right) \qquad (3.1)$$

of occurring. The two forms of (3.1) suggest two methods for attaching the probabilities to the associations. The various associations are effected by

interchanging the rows of the $\underset{\sim}{X}$ matrix. The left-hand form of (3.1) suggests attaching the probability π to every possible interchange of distinct rows of $\underset{\sim}{X}$; the right form suggests attaching the probability $1/N!$ to every possible interchange of rows of $\underset{\sim}{X}$. The lack of distinctness of these row interchanges will supply the multiplier needed for $1/N!$. This latter approach is easier to follow through.

Notation will be simplified if we write $\underset{\sim}{M} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'$, $\underset{\sim}{Q}'\underset{\sim}{Q} = (\underset{\sim}{Y}'\underset{\sim}{Y})^{-}$, where $\underset{\sim}{Q}'$ is $q \times b$, b = rank $(\underset{\sim}{Y})$, and let $\underset{\sim}{Z} = (\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_b) = \underset{\sim}{Y}\underset{\sim}{Q}'$. Then $\underset{\sim}{Z}\underset{\sim}{Z}' = \underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}\underset{\sim}{Y}'$ and $\underset{\sim}{Z}'\underset{\sim}{Z} = (\underset{\sim}{z}_r'\underset{\sim}{z}_s) = \underset{\sim}{I}_b$. (Here, we have assumed that $(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}$ satisfies $(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}\underset{\sim}{Y}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-} = (\underset{\sim}{Y}'\underset{\sim}{Y})^{-}$ as well as $\underset{\sim}{Y}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}\underset{\sim}{Y}'\underset{\sim}{Y} = \underset{\sim}{Y}'\underset{\sim}{Y}$ and that it is symmetric. This class of matrices is not void.) The statistic T can be written in this notation as

$$T = \mathrm{tr}[(\underset{\sim}{Y}'\underset{\sim}{M}\underset{\sim}{Y})(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}] = \mathrm{tr}[\underset{\sim}{M}\underset{\sim}{Z}\underset{\sim}{Z}'] = \sum_{s=1}^{b} \mathrm{tr}[\underset{\sim}{M}\underset{\sim}{z}_s\,\underset{\sim}{z}_s'] = \sum_{s=1}^{b} \underset{\sim}{z}_s'\underset{\sim}{M}\underset{\sim}{z}_s \,. \qquad (3.2)$$

Recall that all variates were corrected for their means. Thus $\underset{\sim}{j}'\underset{\sim}{Z} = \underset{\sim}{j}'\underset{\sim}{Y}\underset{\sim}{Q}' = \underset{\sim}{0}$ and $\underset{\sim}{j}'\underset{\sim}{X} = \underset{\sim}{0}$.

3.1 $E_p(T)$

The previously mentioned row interchanges can be accomplished algebraically by replacing $\underset{\sim}{X}$ in T by $\underset{\sim}{P}\underset{\sim}{X}$ where $\underset{\sim}{P}$ is a permutation matrix, i.e. $T = \sum_s \mathrm{tr}[\underset{\sim}{M}\underset{\sim}{z}_s\underset{\sim}{z}_s']$ changes to

$$T = \sum_s \mathrm{tr}[\underset{\sim}{P}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{P}'\underset{\sim}{z}_s\underset{\sim}{z}_s'] = \sum_s \mathrm{tr}[\underset{\sim}{M}\underset{\sim}{P}'\underset{\sim}{z}_s\underset{\sim}{z}_s'\underset{\sim}{P}] \,. \qquad (3.3)$$

Thus we need only $E_p(\underset{\sim}{z}_s\underset{\sim}{z}_s')$ in order to find $E_p(T)$. Since each diagonal position of $\underset{\sim}{z}_s\underset{\sim}{z}_s'$ will be occupied by a specific z_{si}^2 $(N-1)!$ times in the $N!$ permutations of the rows of $\underset{\sim}{z}_s$,

$$E(\text{a diagonal element of } \underset{\sim}{z}_s\underset{\sim}{z}_s') = \frac{(N-1)!}{N!}\sum_{i=1}^{N} z_{si}^2 = \frac{1}{N}\,\underset{\sim}{z}_s'\underset{\sim}{z}_s = \frac{1}{N}\,.$$

Likewise,

$$E(\text{an off-diagonal element of } \underset{\sim}{z}_s\underset{\sim}{z}_s') = \frac{(N-2)!}{N!}\sum_{i\neq j} z_{si}z_{sj} = \frac{1}{N(N-1)}\left[(\Sigma z_{si})^2 - \Sigma z_{si}^2\right]$$

$$= -\frac{1}{N(N-1)}\,.$$

Combining the above and denoting $\underset{\sim}{j}\underset{\sim}{j}'$ by $\underset{\sim}{J}$, we have

$$E(\underset{\sim}{z}_s\underset{\sim}{z}_s') = \frac{1}{N}\,\underset{\sim}{I}_N - \frac{1}{N(N-1)}\,(\underset{\sim}{J} - \underset{\sim}{I}) = \frac{1}{N-1}\,(\underset{\sim}{I} - \frac{1}{N}\,\underset{\sim}{J})\,,$$

and thus

$$E_p(T) = \sum_s \mathrm{tr}[\underset{\sim}{M}E_p(\underset{\sim}{z}_s\underset{\sim}{z}_s')] = \frac{b}{N-1}\,\mathrm{tr}[\underset{\sim}{M}] - \frac{b}{N(N-1)}\,\mathrm{tr}[\underset{\sim}{M}\underset{\sim}{J}] = \frac{ab}{N-1}\,, \qquad (3.4)$$

where $a = \mathrm{tr}[\underset{\sim}{M}] = \text{rank } (\underset{\sim}{X})$. It should be noted here that in the full rank regression case, $a = p$ and with sufficient observations, $b = q$. When $q = 1$, this result agrees with the univariate formula of Box and Watson [1962].

3.2 $\mathrm{Var}_p(T)$

The first step in deriving the permutation variance of T, $\mathrm{Var}_p(T)$, is to expand it in terms of variances and covariances of the variate components from (3.2):

$$\mathrm{Var}_p(T) = \sum_{s=1}^{b} \mathrm{Var}_p(\underset{\sim}{z}_s'\underset{\sim}{M}\underset{\sim}{z}_s) + \sum_{s\neq t} \mathrm{Cov}_p(\underset{\sim}{z}_s'\underset{\sim}{M}\underset{\sim}{z}_s\,,\ \underset{\sim}{z}_t'\underset{\sim}{M}\underset{\sim}{z}_t)\,. \qquad (3.5)$$

The variance terms in the first sum can be derived simply from results given by

Box and Watson [1962]. They obtained $\mathrm{Var}_p(\underset{\sim}{z}'\underset{\sim}{M}\underset{\sim}{z})$ when $\underset{\sim}{M} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$, i.e. when $\underset{\sim}{X}$ is full rank. They used only the relations $\underset{\sim}{M}\underset{\sim}{j} = \underset{\sim}{0}$, $\underset{\sim}{M}^2 = \underset{\sim}{M}$, and $\mathrm{tr}[\underset{\sim}{M}] = p$ in their argument. Since here we have $\underset{\sim}{M}\underset{\sim}{j} = \underset{\sim}{0}$, $\underset{\sim}{M}^2 = \underset{\sim}{M}$, and $\mathrm{tr}[\underset{\sim}{M}] = a$, their result is applicable to $\mathrm{Var}_p(\underset{\sim}{z}_s'\underset{\sim}{M}\underset{\sim}{z}_s)$ for $s = 1, 2, \cdots, b$ with p replaced by a:

$$\mathrm{Var}_p(\underset{\sim}{z}_s'\underset{\sim}{M}\underset{\sim}{z}_s) = \frac{2a(N-1-a)}{(N+1)(N-1)^2} + k_4^s\left[g - \frac{a(a+2)(N-1)}{N(N+1)}\right]. \tag{3.6}$$

Following through details similar to Box and Watson's, Urquhart [1965] found

$$\mathrm{Cov}_p(\underset{\sim}{z}_s'\underset{\sim}{M}\underset{\sim}{z}_s,\ \underset{\sim}{z}_t'\underset{\sim}{M}\underset{\sim}{z}_t) = -\frac{2a(N-1-a)}{(N+1)(N-1)^2(N-2)} + k_{22}^{st}\left[g - \frac{a(a+2)(N-1)}{N(N+1)}\right]. \tag{3.7}$$

In (3.6) and (3.7), g is the sum of squares of the diagonal elements of $\underset{\sim}{M}$ and k_4^s and k_{22}^{st} are multivariable generalizations of Fisher's k statistics as given by Kendall and Stuart [1963]. Subsequently the bracketed expressions in (3.6) and (3.7) will be denoted by u.

Combining (3.6) and (3.7), we have then

$$\mathrm{Var}_p(T) = \frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)} + u\left[\sum_s k_4^s + \sum_{s\neq t} k_{22}^{st}\right],$$

and setting the bracketed term equal to K gives us

$$\mathrm{Var}_p(T) = \frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)} + uK. \tag{3.8}$$

Except for a multiplier, u and K are the same function of the matrices $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ respectively, i.e. $uK = cf(\underset{\sim}{X})f(\underset{\sim}{Y})$.

4. Approximate Distribution of T

If we define D by

$$\frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)}\, D = uK \ , \tag{4.1}$$

equation (3.8) can be written as

$$\mathrm{Var}_p(T) = (1+D)\, \frac{2ab(N-1-a)(N-1-b)}{(N+1)(N-1)^2(N-2)} \ . \tag{4.2}$$

Two-moment agreement between the permutation distribution of T and the beta distribution

$$\frac{x^{m-1}\,(1-x)^{n-1}}{B(m,n)} \ , \quad 0 \le x \le 1 \ ,$$

is obtained for $m = \delta a$, $n = \delta(N-1-a)$ by setting

$$\delta = \frac{(Nb-2)(N-1) - 2D(N-b-1)}{2(N-1)(N-b-1)(1+D)} \ . \tag{4.3}$$

This beta distribution is the normal theory distribution of T for a or b = 1 and D = 0. For D = 0 and a and b both at least two, this beta distribution is essentially the mathematical approximation to the normal theory distribution of T suggested by Pillai [1955]. The variance form (4.2) and the adjustment factor (4.3) are multivariate generalizations of the corresponding quantities from Box and Watson [1962].

The permutation theory and normal theory distributions will thus be approximated by the same distribution whenever D = 0. This will happen if either u or K is zero. Box and Watson showed that u can be represented as a sum of standardized fourth order k-statistics of the same form as K. If either $\underset{\sim}{X}$ or $\underset{\sim}{Y}$ were normal conditional upon the other, then E(D) = 0.

Thus if either K or u is near zero, the usual normal theory derivation should provide a reasonable approximation to permutation theory. The statistic K can be zero by each element being zero (which is its expected value under normal theory), or the various addends of K can be cancelling positive and negative values. This latter is the same as the net kurtosis of all the variates being zero. Contrast this to univariate analysis where with a single variate an analogous circumstance cannot exist.

If there is reason to suspect that K will not be close to zero, one is still at liberty to choose $\underset{\sim}{X}$ to make $u = 0$. The beta form of Hotelling's T^2 test for the difference between two multivariate means can serve as an illustration of this point. If the first sample mean is based on r_1 observations and the second on r_2, we can take

$$\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{j}_{r_1} & \underset{\sim}{0} \\ \\ \underset{\sim}{0} & \underset{\sim}{j}_{r_2} \end{bmatrix} - \frac{1}{N} \underset{\sim}{j}_N \left(r_1, r_2 \right) = \frac{1}{N} \begin{bmatrix} r_2 \underset{\sim}{j}_{r_1} & -r_2 \underset{\sim}{j}_{r_1} \\ \\ -r_1 \underset{\sim}{j}_{r_2} & r_1 \underset{\sim}{j}_{r_2} \end{bmatrix}$$

The statistic u is then a constant times $N^2+N - 6r_1 r_2$. If r_1 is taken to be larger than r_2, then $N^2+N - 6r_1 r_2 = 0$ when approximately $r_1 = 3.7r_2 - 1$.

Thus the normal theory distribution of T should provide a satisfactory approximation to the permutation distribution of T whenever

i. The distribution of $\underset{\sim}{Y}$ for any $\underset{\sim}{X}$ is normal or

ii. The cumulants of $\underset{\sim}{X}$ for any $\underset{\sim}{Y}$ are normal

or less specifically when

iii. The responses $\underset{\sim}{Y}$ give K = 0 or

iv. The regressors $\underset{\sim}{X}$ give u = 0.

Otherwise the approximate beta obtained from (4.3) should be used.

5. Discussion

Normal-type properties can be introduced into the distribution of the test statistic T either by "nature" through the responses from the sampling units or through the choice of the experimental design. Yates [1964] has endorsed Fisher's view [1935] that the former is the source on which we rely for normality. Nor would one be inclined to look any further when the observed response is an average of responses on several experimental entities drawn randomly from some large population. It is at the other extreme of experimental types, when the units form the whole population to which inference can be made, when they are contiguous in blocks, and when the effect of treatments is a marked stratification of the response, that we feel one may question the normal-derived approximations to the permutation distribution percentiles. This is particularly true and important to consider when treatments convey a strong effect. When the latter is the case, there are three avenues open for solving the inference problem.

(1) One may approximate the distribution of standardized, quadratic, test statistics with the beta distribution where the first two moments agree with those of the permutation distribution. The justification for this rests largely on the study by Pitman [1937] and a scattering of empirical verifications. In more than one hundred sets of real univariate experimental data analyzed by exact permutation methods at the Colorado State University Statistics Laboratory, we have found no case when the normal approximation did not suffice to assess significance. Sample sizes ranged from six to one hundred and designs ranged from paired t comparisons to replicated randomized blocks. Differences between significance probabilities calculated by the normal theory derivations and from the permutation tests seldom differed by more than 0.01 and then only for small sample sizes where the discreteness in the permutation distribution should have been considered. The only

extension for a multivariate response (a paired Hotelling's T^2 test) which was investigated up to four moments by Arnold [1964], likewise appears to justify the procedure.

(2) One may evaluate the significance of the observed outcome from the permutation distribution. Complete enumeration is possible only for very small numbers of experimental units. Monte Carlo approximations will handle somewhat larger types of experiments, but really large numbers and many restrictions appear at the present to be out of hand for enumeration or sampling techniques.

(3) For test statistics in which the only permutation-variable elements arise from the product $\underset{\sim}{Y}'\underset{\sim}{X}$, a permutation of the association, $\underset{\sim}{Y}'\underset{\sim}{P}\underset{\sim}{X}$, is achieved either by permuting the rows of $\underset{\sim}{Y}$ or the rows of $\underset{\sim}{X}$. The rows of $\underset{\sim}{X}$ can be chosen to approximate a normal distribution to a given degree in a discrete manner. When there are a distinct row vectors in $\underset{\sim}{X}$ then one may choose these to fit the cumulants or k-statistics of $\underset{\sim}{X}$ up to the a^{th} order with those of the normal. This method works very well when p is small, say 2, and the range of the rows of $\underset{\sim}{X}$ is not restricted. It does so however at the expense of efficient use of estimation resources when the mean response has a known polynomial form. It may be possible to use this approach in experiments of blocked treatment plans. If the loss in sensitivity compared to a well balanced design is nominal, such designs would be valuable in those situations where a fine degree of approximation to the distribution of the design responses is needed. We are now investigating these problems.

In the realm of questions about a good test statistic for the multivariate analysis of variance we are inclined to rely on the leading statement in (3) to justify the choice of T. The statistics which have so far received considerable attention,

$$\lambda = | \underset{\sim}{I} - \underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-} |$$

$$T_0^2 = \mathrm{tr}\left\{[\underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}]\,[\underset{\sim}{Y}'(\underset{\sim}{I}-\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}')\underset{\sim}{Y}]^{-}\right\}$$

$$c = \max \mathrm{ch}\left\{[\underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}]\,[\underset{\sim}{Y}'(\underset{\sim}{I}-\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}')\underset{\sim}{Y}]^{-}\right\}$$

$$T = \mathrm{tr}[\underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-}]$$

all vary under randomization according to the manner in which the product $\underset{\sim}{Y}'\underset{\sim}{X}$ varies. Because $\underset{\sim}{Y}'\underset{\sim}{Y}$ and $\underset{\sim}{X}'\underset{\sim}{X}$ do not vary, e.g. $\underset{\sim}{Y}'\underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{Y} = \underset{\sim}{Y}'\underset{\sim}{Y}$, the elements of the generalized inverse matrices in T and λ do not vary while those in T_0^2 and c do. The point in favor of T is that it is the only statistic which is a standardized quadratic in the elements of $\underset{\sim}{X}'\underset{\sim}{Y}$. If either through the rows of $\underset{\sim}{X}$ or the rows of $\underset{\sim}{Y}$ the first a k-statistics of $\underset{\sim}{Y}'\underset{\sim}{X}$ agree closely with those of the normal, then to a corresponding degree of approximation those of T fit the normal-derived distribution of T. This closeness of agreement, available only in statistics which vary as standardized quadratics, should give T an advantage in robustness over λ, T_0^2, and c.

References

Arnold, H. J. (1964). Permutation support for multivariate techniques. Biometrika, 51, 65-70.

Box, G. E. P. and Watson, G. S. (1962). Robustness to non-normality of regression tests. Biometrika, 49, 93-106.

Fisher, R. A. (1935, 1960). The Design of Experiments. First and Seventh Editions. Edinburgh: Oliver and Boyd.

Kendall, M. G. and Stuart, A. (1963). The Advanced Theory of Statistics (Vol. 1). Second Edition. London: Griffin and Company.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. Ann. Math. Statist., 26, 117-121.

Pillai, K. C. S. and Mijares, T. A. (1959). On the moments of the trace of a matrix and approximations to its distribution. Ann. Math. Statist., 30, 1135-1140.

Pitman, E. J. G. (1937). Significance tests which can be applied to samples from any population III. The analysis of variance test. Biometrika, 29, 322-335.

Searle, S. R. (1966). Matrix Algebra for the Biological Sciences. New York: John Wiley and Sons.

Urquhart, N. S. (1965). On the permutation distribution of a multivariate test statistic. Mimeo. Series No. 1, Department of Mathematics and Statistics, Colorado State University, Fort Collins.

Welch, B. L. (1937). On the z-test in randomised blocks and latin squares. Biometrika, 29, 21-52.

Yates, F. (1964). Sir Ronald Fisher and the design of experiments. Biometrics, 20, 307-321.